Supplementary Information to:

Annotation of loci from genome-wide association studies using tissue-specific quantitative interaction proteomics

Alicia Lundby*, Elizabeth J. Rossin*, Annette B. Steffensen, Moshe Rav Acha, Christopher Newton-Cheh, Arne Pfeufer, Stacey N. Lynch, The QT-IGC Consortium, Søren-Peter Olesen, Søren Brunak, Patrick T. Ellinor, J.Wouter Jukema, Stella Trompet, Ian Ford, Peter W. Macfarlane, Bouwe P. Krijthe, Albert Hofman, Andre G. Uitterlinden, Bruno H. Stricker, Hendrick M. Nathoe, Wilko Spiering, Mark J. Daly, Folkert W. Asselbergs, Pim van der Harst, David J. Milan, Paul I. W. de Bakker#, Kasper Lage#, and Jesper V. Olsen#.

*These authors contributed equally.

#These authors also contributed equally.

Correspondence: lage.kasper@mgh.harvard.edu or jesper.olsen@cpr.ku.dk

Supplementary Figure 1.

Amino acid sequence coverage and representative spectra for KCNQ1.

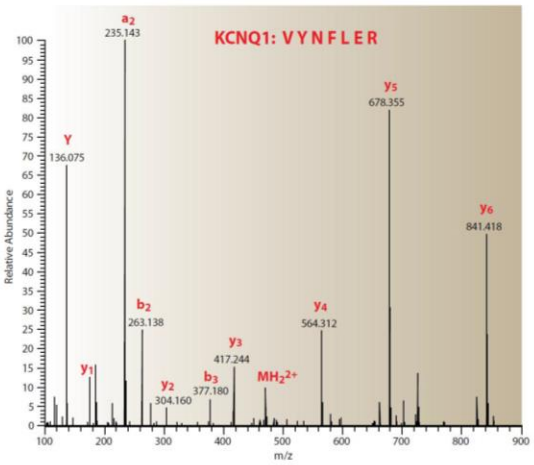
KCNQ1_MOUSE - LQT1 protein:

Potassium voltage-gated channel subfamily KQT member 1 (Voltage-gated potassium channel subunit Kv7.1) (IKs producing slow voltage-gated potassium channel subunit alpha KvLQT1) (KQT-like 1) - Mus musculus (Mouse)

Matched peptides shown in **Bold Red**

```
  1 MDTASSPPSA ERKRAGWSRL LGARRGSAVV KKCPFSLELA EGGPEGSTVY
 51 APIAPTGAPG LAPPMSTPVS PAPAPADLGP RPRVSLDPRV SIYSARRPLL
101 ARTHIQGRVY NFLERPTGWK CFVYHFTVFL IVLVCLIFSV LSTIEQYAAL
151 ATGTLFWMEI VLVVFFGTEY VVRLWSAGCR SKYVGIWGRL RFARKPISII
201 DLIVVVASMV VLCVGSKGQV FATSAIRGIR FLQILRMLHV DRQGGTWRLL
251 GSVVFIHRQE LITTLYIGFL GLIFSSYFVY LAEKDAVNES GRIEFGSYAD
301 ALWWGVVTVT TIGYGDKVPQ TWVGKTIASC FSVFAISFFA LPAGILGSGF
351 ALKVQQKQRQ KHFNRQIPAA ASLIQTAWRC YAAENPDSAT WKIYVRKPAR
401 SHTLLSPSPK PKKSVMVKKK KFKLDKDNGM SPGEKMFNVP HITYDPPEDR
451 RPDHFSIDGY DSSVRKSPTL LELSTPHFLR TNSFAEDLDL EGETLLTPIT
501 HVSQLRDHHR ATIKVIRRMQ YFVAKKKFQQ ARKPYDVRDV IEQYSQGHLN
551 LMVRIKELQR RLDQSIGKPS LFIPISEKSK DRGSNTIGAR LNRVEDKVTQ
601 LDQRLVIITD MLHQLLSMQQ GGPTCNSRSQ VVASNEGGSI NPELFLPSNS
651 LPTYEQLTVP QTGPDEGS
```

MS/MS spectrum of KCNQ1 peptide:



Peptides identified by MS analysis are shown in red, peptides not identified in MS analyses are black, and blue boxes indicate membrane spanning regions that are unlikely to be captured by MS analysis for KCNQ1. The figure illustrates that MS analysis provides extensive coverage of the accessible sequences. Below the amino acid sequence we show a representative MS/MS spectrum, labeling the ions detected.

Supplementary Figure 2.

Amino acid sequence coverage and representative spectra for KCNH2

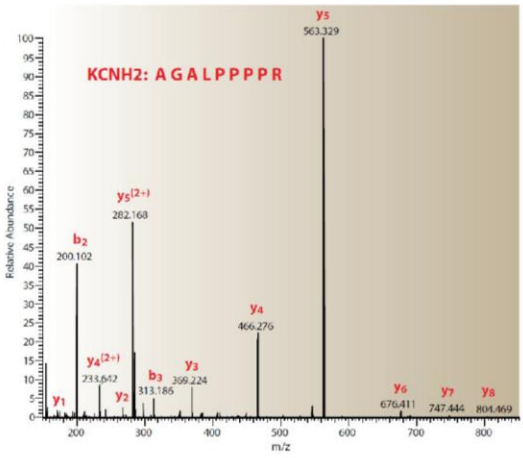
KCNH2_MOUSE - LQT2 protein:

Isoform 1 of Potassium voltage-gated channel subfamily H member 2
Ether-a-go-go-related gene potassium channel 1 (Erg-1)

Matched peptides shown in **Bold Red**

```
   1 MPVRRGHVAP QNTFLDTIIR KFEGQSRKFI IANARVENCA VIYCNDGFCE
  51 LCGYSRAEVM QRPCTCDFLH GPRTQRRAAA QIAQALLGAE ERKVEIAFYR
 101 KDGSCFLCLV DVVPVKNEDG AVIMFILNFE VVMEKDMVGS PAHDTNHRGP
 151 STSWLASGRA KTFRLKLPAL LALTARESSV RTGSMHSAGA PGAVVVDVDL
 201 TPAAPSSESL ALDEVSAMDN HVAGLGPAEE RRALVGPGSA SPVASIRGPH
 251 PSPRAQSLNP DASGSSCSLA RTRSRESCAS VRRASSADDI EAMRAGALPP
 301 PPRHASTGAM HFLRSGLLNS TSDSDLVRYR TISKIPQITL NFVDLKGDPF
 351 LASPTSDREI IAPKIKERTH NVTEKVTQVL SLGADVLPEY KLQAPRIHRW
 401 TILHYSPFKA VWDWLILLLV IYTAVFTPYS AAFLLKETED GSQAPDCGYA
 451 CQPLAVVDLI VDIMFIVDIL INFRTTYVNA NEEVVSHPGR IAVHYFKGWF
 501 LIDMVAAIPF DLLIFGSGSE ELIGLLKTAR LLRLVRVARK LDRYSEYGAA
 551 VLFLLMCTFA LIAHWLACIW YAIGNMEQPH MDSHIGWLHN LGDQIGKPYN
 601 SSGLGGPSIK DKYVTALYFT FSSLTSVGFG NVSPNTNSEK IFSICVMLIG
 651 SLMYASIFGN VSAIIQRLYS GTARYHTQML RVREFIRFHQ IPNPLRQRLE
 701 EYFQHAWSYT NGIDMNAVLK GFPECLQADI CLHLNRSLLQ HCKPFRGATK
 751 GCLRALAMKF KTTHAPPGDT LVHAGDLLTA LYFISRGSIE ILRGDVVVAI
 801 LGKNDIFGEP LNLYARPGKS NGDVRALTYC DLHKIHRDDL LEVLDMYPEF
 851 SDHFWSSLEI TFNLRDTNMI PGSPGSAELE SGFNRQRKRK LSFRRRTDKD
 901 TEQPGEVSAL GQGPARVGPG PSCRGQPGGP WGESPSSGPS SPESSEDEGP
 951 GRSSSPLRLV PFSSPRPPGD PPGGEPLTED GEKSDTCNPL SGAFSGVSNI
1001 FSFWGDSRGR QYQELPRCPA PAPSLLNIPL SSPGRRSRGD VESRLDALQR
1051 QLNRLETRLS ADMATVLQLL QRQMTLVPPA YSAVTTPGPG PTSASPLLPV
1101 GPVPTLTLDS LSQVSQFVAF EELPAGAPEL PQDGPTRRLS LPGQLGALTS
1151 QPLHRHGSDP GS
```

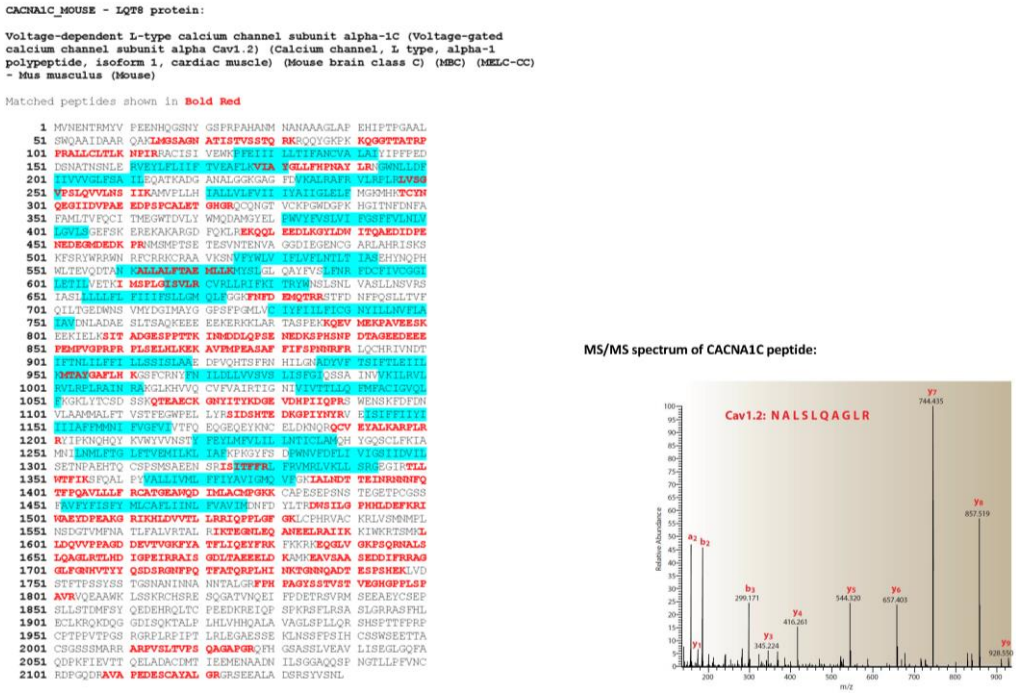
MS/MS spectrum of KCNH2 peptide:



Peptides identified by MS analysis are shown in red, peptides not identified in MS analyses are black, and blue boxes indicate membrane-spanning regions that are unlikely to be captured by MS analysis for KCNH2.

Supplementary Figure 3.

Amino acid sequence coverage and representative spectra for CACNA1C



Peptides identified by MS analysis are shown in red, peptides not identified in MS analyses are black, and blue boxes indicate membrane spanning regions that are unlikely to be captured by MS analysis for CACNA1C. The figure illustrates that MS analysis provides extensive coverage of the accessible sequences. Below the amino acid sequence we show a representative MS/MS spectrum, labeling the ions detected.

Supplementary Figure 4.

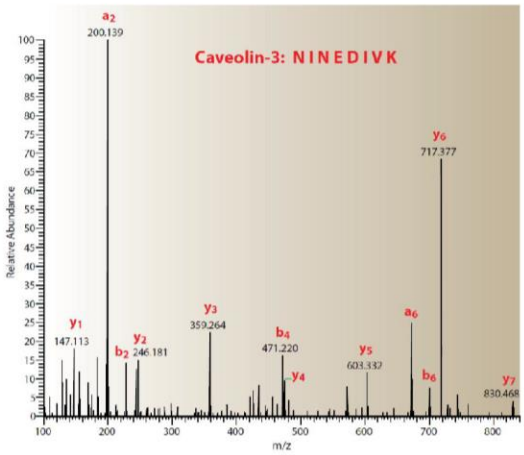
Amino acid sequence coverage and representative spectra for CAV3.

CAV3_MOUSE - LQT11 protein:

Caveolin-3 (M-caveolin) - Mus musculus (Mouse)
Matched peptides shown in **Bold Red**

```
  1 MMTEEHTDLE ARIIKDIHCK EIDLVNRDPK NINEDIVKVD FEDVIAEPEG
 51 TYSFDGVWKV SFTTFTVSKY WCYRLLSTLL GVPLALLWGF LFACISFCHI
101 WAVVPCIKSY LIEIQCISHI YSLCIRTFCN PLFAALGQVC SNIKVVLRRE
151 G
```

MS/MS spectrum of CAV3 peptide:



Match to: **CAV1_MOUSE**
Caveolin-1 - Mus musculus (Mouse)
Matched peptides shown in **Bold Red**

```
  1 MSGGKYVDSE GHLYTVPIRE QGNIYKPNNK AMADEVTEKQ VYDAHTKEID
 51 LVNRDPKHLN DDVVKIDFED VIAEPEGTHS FDGIWKASFT TFTVTKYWFY
101 RLLSTIFGIP MALIWGIYFA ILSFLHIWAV VPCIKSFLIE IQCISRVYSI
151 YVHTFCDPLF EAIGKIFSNI RISTQKEI
```

Match to: **CAV2_MOUSE**
Caveolin-2 - Mus musculus (Mouse)
Matched peptides shown in **Bold Red**

```
  1 MGLETEKADV QLFMADDAYS HHSGVDYADP EKYVDSSHDR DPHQLNSHLK
 51 LGFEDLIAEP ETTHSFDKVW ICSHALFEIS KYVMYKFLTV FLAIPLAFIA
101 GILFATLSCL HIWILMPFVK TCLMVLPSVQ TIWKSVTDVV IGPLCTSVGR
151 SFSSVSMQLS HD
```

The amino acid sequences of CAV3 as well as of CAV1 and CAV2 are shown to illustrate that we can specifically identify CAV3. Peptides identified by MS analysis are shown in red, peptides not identified in MS analyses are black, and blue boxes indicate membrane spanning regions that are unlikely to be captured by MS analysis. Below the amino acid sequence we show a representative MS/MS spectrum, labeling the ions detected.

Supplementary Figure 5.

Amino acid sequence coverage and representative spectra for SNTA1.

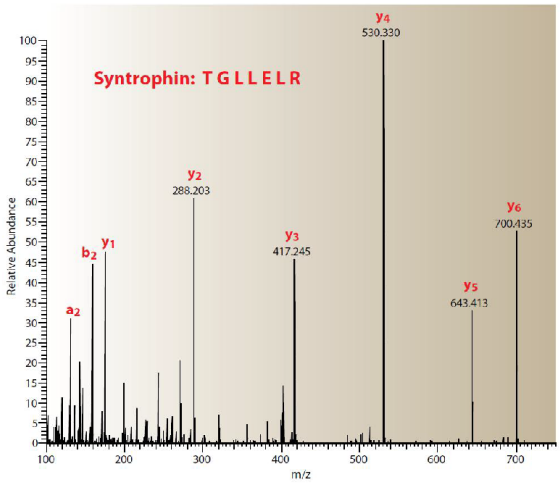
SNTA1_MOUSE - LQT12 protein:

Alpha-1-syntrophin (59 kDa dystrophin-associated protein A1 acidic component 1) (Syntrophin 1) - Mus musculus (Mouse)

Matched peptides shown in **Bold Red**

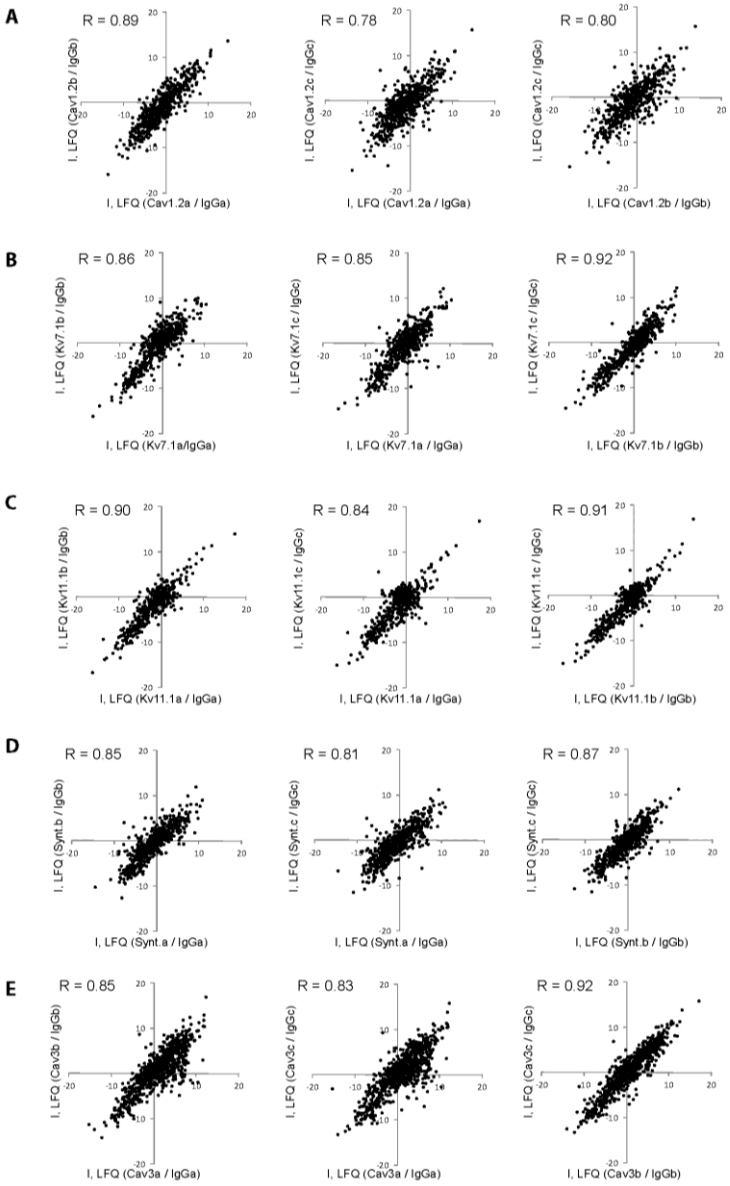
```
   1 MASGRRAPRT GLLELRCGAG SGAGGERWQR VLLSLAEDAL TVSPADGEPG
  51 PEPEPAQLNG AAEPGAAPPQ LPEALLLQRR RVTVRKADAG GLGISIKGGR
 101 ENKMPILISK IFKGLAADQT EALFVGDAIL SVNGEDLSSA THDEAVQALK
 151 KTGKEVVLEV KYMKEVSPYF KNSAGGTSVG WDSPPASPLQ RQPSSPGPQP
 201 RNLSEAKHVS LKMAYVSRRC TPTDPEPRYL EICAADGQDA VFLRAKDEAS
 251 ARSWAGAIQA QIGTFIPWVK DELQALLTAT GTAGSQDIKQ IGWLTEQLPS
 301 GGTAPTLALL TEKELLFYCS LPQSREALSR PTRTAPLIAT RSAHRLVHSG
 351 PSKGSVPYDA ELSFALRTGT RHGVDTHLFS VESPQELAAW TRQLVDGCHR
 401 AAEGIQEVST ACTWNGRPCS LSVHIDKGFT LWAAEPGAAR AMLLRQPFEK
 451 LQMSSDDGTS LLFLDFGGAE GEIQLDLHSC PKTMVFIIHS FLSAKVTRLG
 501 LLA
```

MS/MS spectrum of SNTA1 peptide:



The amino acid sequence of SNTA1 is shown above. Peptides identified by MS analysis are shown in red, peptides not identified in MS analyses are black, and blue boxes indicate membrane-spanning regions that are unlikely to be captured by MS analysis. The figure illustrates that MS analysis provides extensive coverage of the accessible sequences. Below the amino acid sequence we show a representative MS/MS spectrum, labeling the ions detected.

Supplementary Figure 6. Reproducibility of bait immunoprecipitations.



For each bait, triplicates of IPs were performed (experiments a, b, c). Ratios of $\log_{10}$-transformed label-free quantified (LFQ) intensities of the identified proteins were calculated for bait IPs relative to IgG control IPs and plotted against each other for the three triplicate experiments, CACNA1C (also known as Cav1.2) (a), KCNQ1 (also known as Kv7.1) (b), KCNH2 (also known as Kv11.1) (c), SNTA1 (also known as Snta1) (d), CAV3 (also known as Cav3) (e). In each plot the Pearson correlation coefficient is indicated in the upper left corner.

Supplementary Figure 7. The robustness of using IgGs as control.

	CM1-5_average	CM1-5_median	CM1
#Cav1.2 interactors identified (IgGs as controls): 85			
# proteins that overlap with IgG controls	76	76	63
Percent of specific interactors identified	0.89	0.89	0.74

	CM1-5_average	CM1-5_median	CM1
#Kv7.1 interactors identified (IgGs as controls): 116			
# proteins that overlap with IgG controls	98	105	83
Percent of specific interactors identified	0.84	0.91	0.72

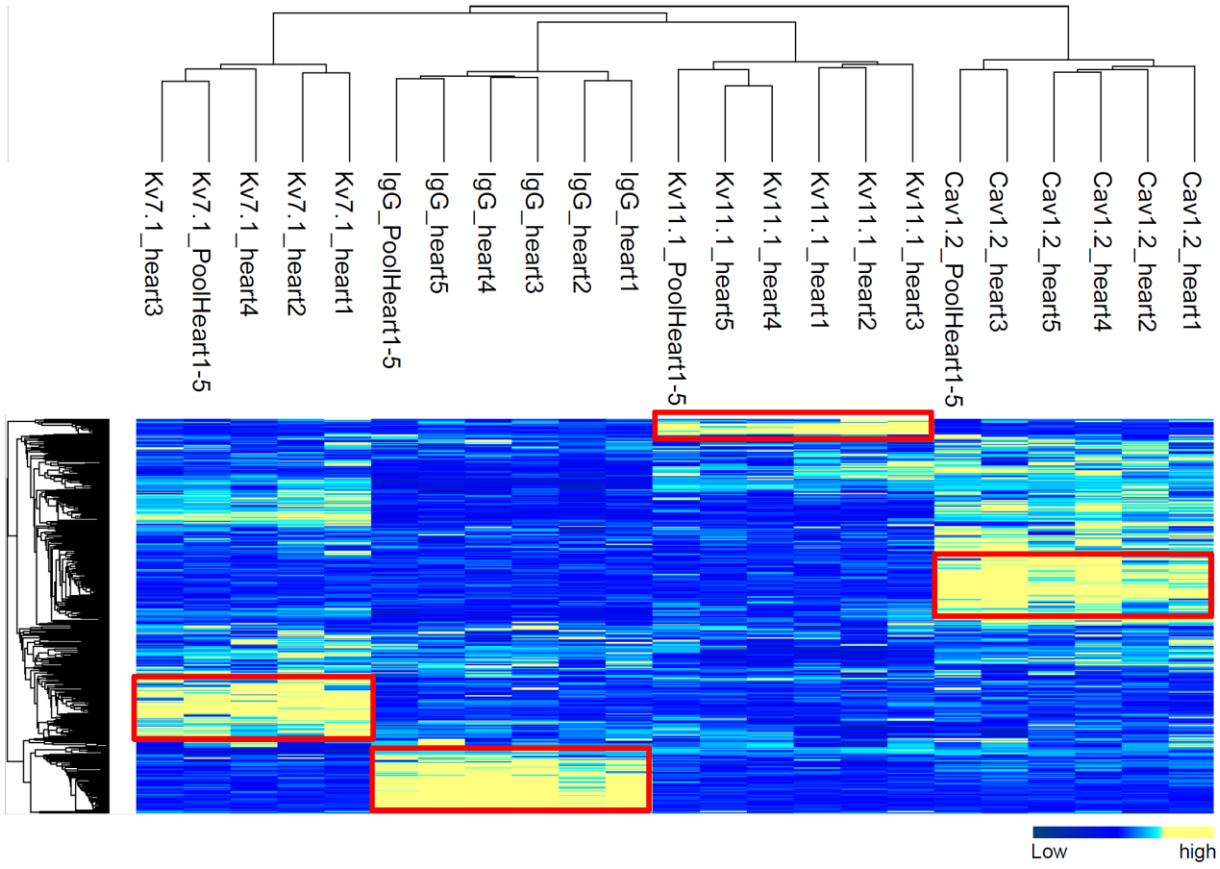
	CM1-5_average	CM1-5_median	CM1
#Kv11.1 interactors identified (IgGs as controls): 31			
# proteins that overlap with IgG controls	27	30	21
Percent of specific interactors identified	0.87	0.97	0.68

	CM1-5_average	CM1-5_median	CM1
#Snta1 interactors identified (IgGs as controls): 104			
# proteins that overlap with IgG controls	86	90	84
Percent of specific interactors identified	0.83	0.87	0.81

	CM1-5_average	CM1-5_median	CM2
#Cav3 interactors identified (IgGs as controls): 333			
# proteins that overlap with IgG controls	301	304	302
Percent of specific interactors identified	0.90	0.91	0.91

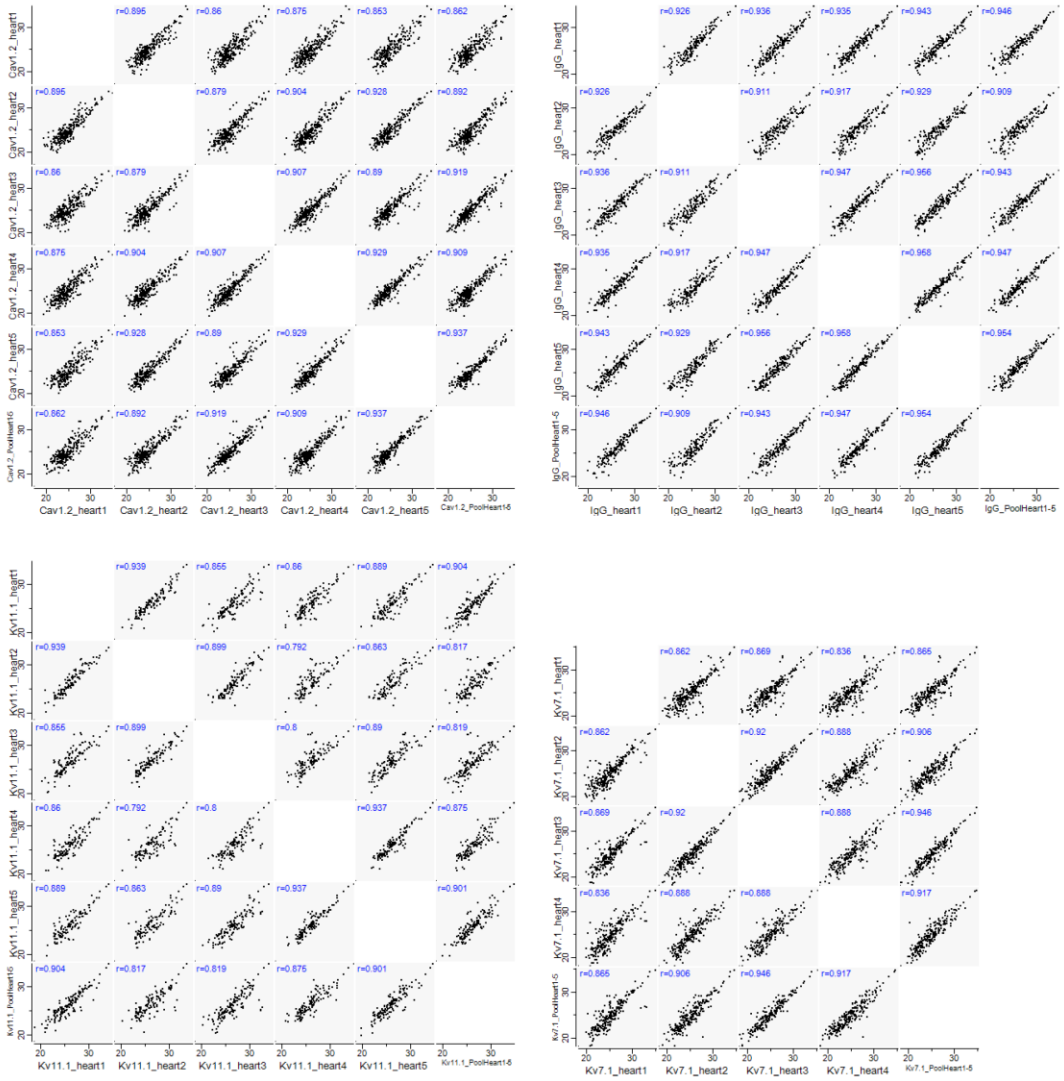
Three different control procedures (not based on IgGs) were tested focusing on five heart proteins (named CM1_5 for cardiomyopathy protein 1 to 5). We made IPs of CM1_5 and analyzed their interaction partners by a similar approach as used for the LQTS proteins. Similar to how we analyzed the LQTS IPs versus IgG IPs, we here analyzed the LQTS IPs versus i) the average protein intensities for CM1_5 (CM1_5-average), ii) the median protein intensities for CM1_5 (CM1_5-median), or iii) the heart interactome that the particular LQTS pulldown is most similar to (which was either CM1 or CM2). There is a high degree of consistency between the proteins interacting with each of the LQTS proteins (KCNQ1 also known as Kv7.1; KCNH2 also known as Kv11.1; CACNA1C also known as Cav1.2; Cav3; and Snta1), regardless of the control procedure, showing the robustness of using an IgG control. Using the median of all 5 cardiomyopathy pull-downs as the control, we identify between 87% and 97% (average 91%) of the interaction partners identified with the IgG control procedure. Using the average of all 5 cardiomyopathy pull-downs as the control, we identify between 83% and 90% (average 87%) of the interaction partners identified using the IgGs as the control. Testing each of the LQTS pull-downs against the most similar cardiomyopathy pull-down, we identify between 68% and 91% (average 77%) of the same interaction partners identified using the IgG control procedure.

Supplementary Figure 8. Clustering analysis of all IPs performed in this study.



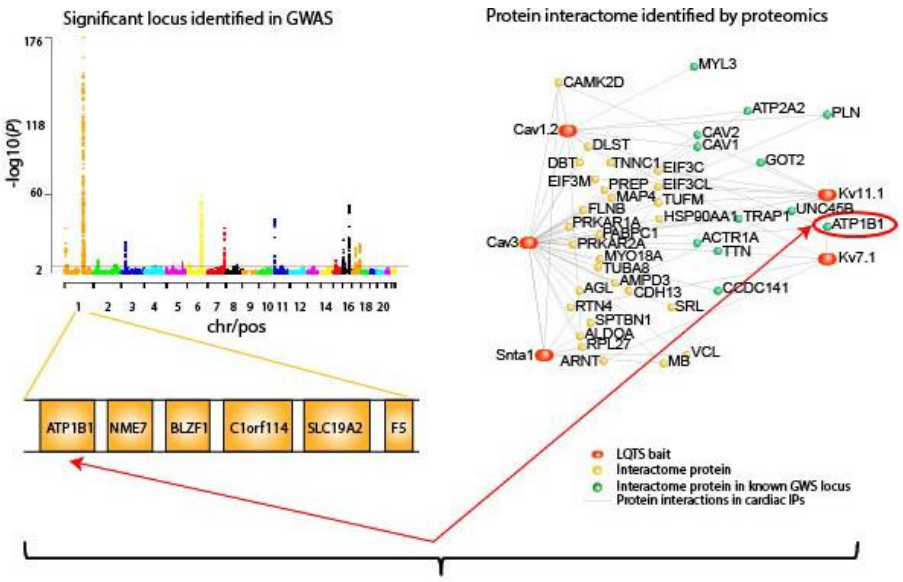
The heat map clearly shows that the IPs targeting the same bait (KCNQ1 [also known as Kv7.1]; KCNH2 [also known as Kv11.1]) cluster when considered across different hearts, the IPs from pooled heart samples cluster with the analogous IPs from the individual hearts. These results show that the interaction partners we identify with the different baits using technical replicates (pooled hearts), are highly similar to the interaction partners identified using biological replicates (hearts 1-5).

Supplementary Figure 9 Correlation plots LFQ intensities for the four sets of IPs (KCNQ1 [also known as Kv7.1], KCNH2 [also known as Kv11.1], CACNA1C [also known as Cav1.2] and IgGs).



For each plot the Pearson correlation coefficient is provided in the upper left corner. The average correlation coefficient between a pooled heart sample and the individual heart samples is 0.91 (or 0.93 for CACNA1C; 0.94 for IgG; 0.86 for KCNH2; and 0.91 for KCNQ1).

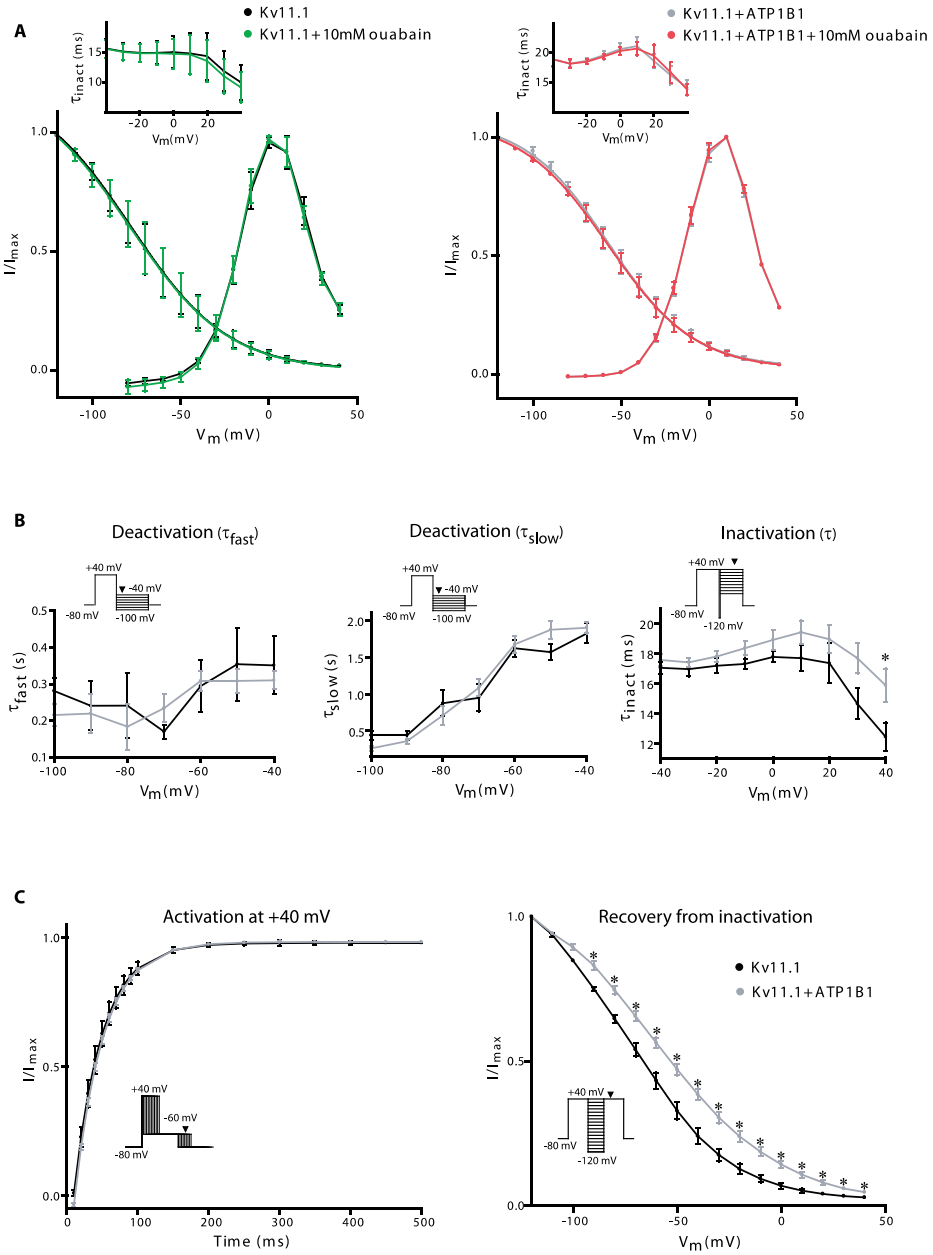
Supplementary Figure 10. Quantitative interaction proteomics annotate loci identified in GWA studies.



SNP	#genes in locus	Genes from locus identified in interactomes
rs3026445	21	ATP2A2
rs10919070	6	ATP1B1
rs9920	2	CAV1, CAV2
rs7561149	4	CCDC141, TTN
rs457162	3	PLN
rs1634800	8	UNC45B
rs1296720	3	TRAP1
rs1549607	5	GOT2
rs2485376	14	ACTR1A
rs17784882	13	MYL3

As exemplified by the GWAS identified locus encompassing ATP1B1 along with 5 other genes, a locus identified in GWAS often contains multiple genes. Therefore functional biological data of relevance to the GWA study is necessary in order to pinpoint the specific gene causing the phenotype. From our proteomic data we identified the protein encoded by ATP1B1 in the protein networks of the Mendelian LQTS proteins KCNH2 (also known as Kv11.1), KCNQ1 (also known as Kv7.1), CACNA1C (also known as Cav1.2), SNTA1, and CAV3 thus providing functional data that ATP1B1 is the likely causal gene associated with the phenotype in the locus. As highlighted in the table, our experimental data suggest loci annotation in eight cases. In the table we indicate the number of genes residing in the respective GWAS identified loci as well as the genes represented in the networks.

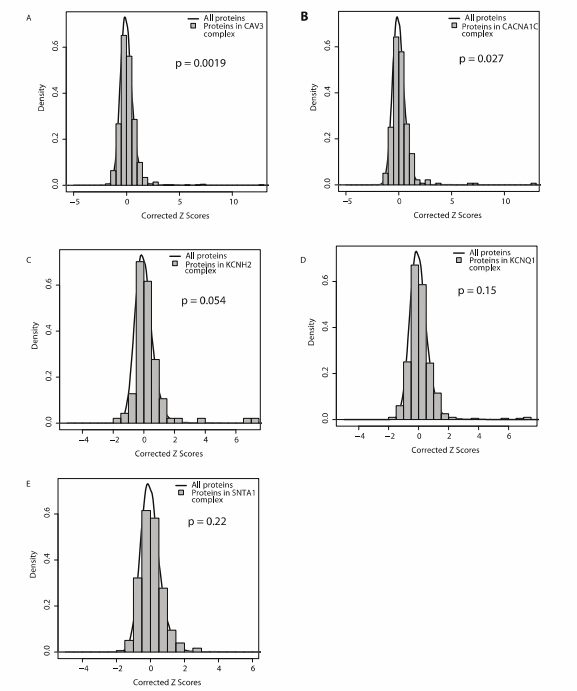
Supplementary Figure 11. Electrophysiological measurements of KCNH2 (also known as Kv11.1) or KCNH2+Atp1b1 from *Xenopus laevis* oocytes.



(A) To ascertain that the effect observed upon co-expression with Atp1b1 was a direct effect on the KCNH2 channel, and not due to Atp1b1 affecting endogenous Atp1a1 generated pump current, we made a two-fold test. First, we tested currents generated by *Xenopus laevis* oocytes only expressing Atp1b1. These oocytes conducted currents of very low magnitude, comparable to the endogenous currents recorded from oocytes injected with water (data not shown).

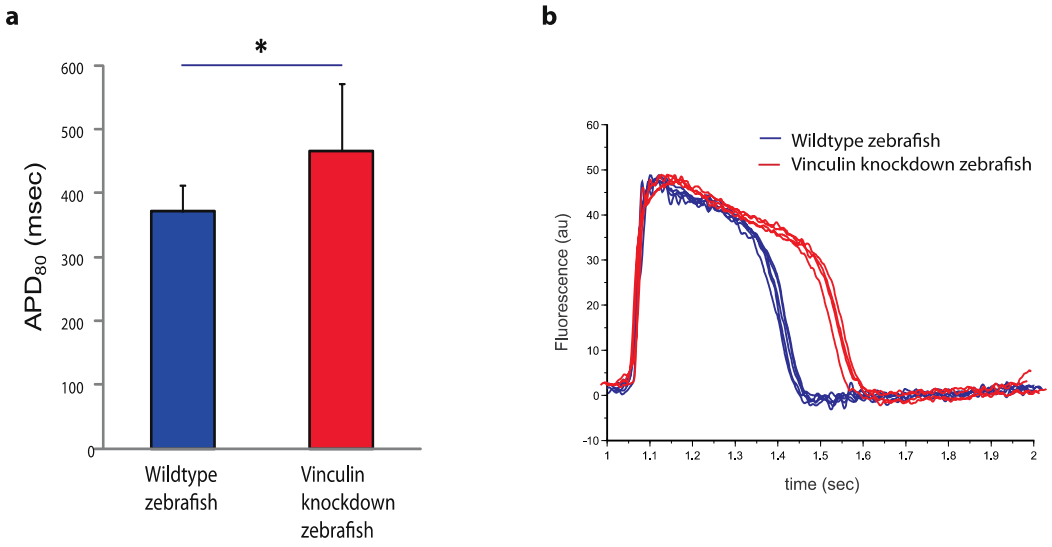
Second, to further confirm that the current effect is not due to ATP1B1 regulating endogenous ATP1A1 we tested the effect on KCNH2 or KCNH2+ ATP1B1 generated currents in the presence of 10mM ouabain. Ouabain is a well characterized inhibitor of Atp1a1, and ouabain did not affect the currents generated by neither KCNH2 (n=3) nor KCNH2+ ATP1B1 (n=3). The I-V curve, the recovery from inactivation as well as inactivation time-constants are shown for KCNH2 (left) with and without 10mM ouabain, and the same is shown for KCNH2+ ATP1B1 (right) with and without 10mM ouabain. (B) Channel kinetics were evaluated as a function of the membrane potential by calculating fast and slow deactivation time-constants as well as the inactivation time constants from recordings elicited by the depicted clamp protocols. Triangles indicate the points of the clamp protocols used for the analysis. The kinetics of deactivation were determined by recording the tail currents at potentials ranging from -40 to -100 mV in 10mV decrements after an activation step to +40 mV and fitting the currents to a double-exponential function. The time constants were plotted against the membrane potential and neither the fast nor the slow component were affected by ATP1B1 co-expression, KCNH2 (n=9) and KCNH2+ ATP1B1 (n=10). The relative contribution of the two deactivation components were analysed from their respective amplitudes and again no variations were found (data not shown). To investigate inactivation kinetics the channels were first fully inactivated at +40 mV, released from inactivation by a brief hyperpolarizing step to -120mV, followed by a final step to potentials ranging from +40 to -40 mV in 10mV decrements, where tail currents were measured. Inactivation time-constants were evaluated from mono-exponential fits to the tail currents, and were plotted against the applied voltage. (C) Channel activation as well as recovery from inactivation was evaluated for KCNH2 or KCNH2+ ATP1B1 channels based on recordings using the clamp protocols shown as insets, where triangles indicate measurement points used for analysis. For channel activation, an evelope-of-tails protocol was applied: channels were activated at +40 mV for various durations (10-500 ms) followed by a hyperpolarization to -60 mV, where the peak tail current was measured. The peak tail amplitudes were normalized to the maximum amplitude and plotted as a function of time and time constants were determined by fitting the data to mono-exponential functions. For KCNH2 (n=7) τ=40.2±4.9 ms and for KCNH2+ ATP1B1 (n=10) τ=40.7±1.5 ms. The voltage dependent recovery from inactivation was investigated by a 3 step protocol: the potential was clamped at +40mV to assure full inactivation of the channels, then the channels were subjected to a series of brief (10ms) hyperpolarizing steps from +40 to -120 mV in 10mV decrements, followed by a return to +40mV. The peak current amplitude after return to +40 mV was measured and normalized to the maximum current level and plotted against the potential at the hyperpolarized step. Data were fitted to a Boltzmann equation. For KCNH2 (n=7) V½=-74.5±2 mV and for KCNH2+ ATP1B1 (n=11) V½= -60.7±1.8 mV.

Supplementary Figure 12. Association analysis for the individual networks. CAV3 (a), CACNA1C (b), KCNH2 (c), KCNQ1 (d), SNTA1 (e).



We used data on SNPs associated to QT interval variation from the QT-IGC meta-analysis and replication in >100,000 individuals of European ancestry. Association Z-scores were derived for individual genes as described above under "methods", and we depict the distribution of the association Z-scores for genes represented in the individual networks (grey bars) to a background distribution of all genes in the genome (black line). The x-axis represents Z-scores assigned to genes corrected for SNP density and linkage disequilibrium structure. Despite the reduced statistical power associated with lowering the proteomics coverage, two of the five LQTS networks are enriched in proteins encoded on regions significantly associated to common QT interval variation (CAV3 P = 2.27e-4, CACNA1C P = 0.0017) and via a composite test of genetic association one is individually associated (CAV3 P = 0.0019).

Supplementary Figure 13. *Vinculin* knockdown prolongs action potential duration in zebrafish.



(A) Morpholino knockdown of zebrafish *vinculin* resulted in prolonged cardiac action potentials (APD_{80} =466±105 msec) compared to carrier injected controls (APD_{80} = 371±40 msec), *P* = 0.04, n = 13 independent samples. Superimposed exemplar traces are shown for one representative sample for *Vinculin* knockdown (red) and Control conditions (blue).

Reference list:

15. Lage, K. *et al.* A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 20870–20875 (2008).
16. Lage, K. *et al.* A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat. Biotechnol.* **25,** 309–316 (2007).
26. Gavin, A.-C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440,** 631–636 (2006).
37. Krogan, N. J. *et al.* Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature* **440,** 637–643 (2006).